

Challenges for the estimation of uncertainty of measurements made in situ

Article (Published Version)

Ramsey, Michael (2020) Challenges for the estimation of uncertainty of measurements made in situ. Accreditation and Quality Assurance. ISSN 0949-1775

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/93949/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Challenges for the estimation of uncertainty of measurements made in situ

Michael H. Ramsey

Accreditation and Quality Assurance

Journal for Quality, Comparability and
Reliability in Chemical Measurement

ISSN 0949-1775

Accred Qual Assur

DOI 10.1007/s00769-020-01446-4



Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.



Challenges for the estimation of uncertainty of measurements made in situ

Michael H. Ramsey¹

Received: 17 March 2020 / Accepted: 4 September 2020
© The Author(s) 2020

Abstract

In situ measurements are made without the removal of a physical sample and have many advantages over traditional ex situ measurements, made on a removed sample usually in a remote laboratory. The quality of ex situ measurements is usually expressed primarily in terms of their measurement uncertainty, including that arising during the sampling process. However, estimates of uncertainty for in situ measurement values have not usually included this uncertainty from sampling (U_fS). It is argued that the making of an in situ measurement inevitably includes the taking of an ‘undisturbed sample’ that generates U_fS, which should be included in the estimate of measurement uncertainty. Because undisturbed samples are not prepared or mixed, as is usual for removed samples, the heterogeneity of the analyte concentration in the sampling target is the primary source of U_fS. Existing methods for estimating U_fS for ex situ measurements can broadly be applied to in situ measurements. However, four extra challenges that limit the design and uptake of uncertainty estimation for in situ methods are identified, and possible solutions and actions required are discussed. Examples of in situ measurements considered include Pb in top soil by hand-held PXRF, ¹³⁷Cs at a nuclear site by portable gamma-ray spectrometry, and bilirubin in new-born infants by hand-held reflectance photometry.

Keywords In situ measurement · Uncertainty · Sampling · Heterogeneity

Introduction

This paper aims to describe in situ measurements in general, identify the challenges there are in estimating their uncertainty, and suggest possible solutions to these challenges. In situ measurements of chemical concentration are made at the original location of the test material without the removal of a physical sample. Such measurements are now becoming more prevalent than traditional ex situ measurements. In situ measurements cover an enormous diversity of analytes, targets and situations at a wide range of different measurement scales, from the macro (e.g. centimetre scale with PXRF [1]) down to the micro (e.g. microns with SIMS [2]), with

the mass of the corresponding ‘undisturbed sample’ ranging from micrograms to picograms.

It is widely accepted that when such measurements are made ex situ in a laboratory, be that remote or ‘on site’, then the analytical method needs to be validated. The key metric of the quality of any measurement value is its uncertainty, which can be used to judge whether a measurement is fit for its intended purpose (FFP). However, this rigorous approach has not been widely applied to measurements that are made in situ, for reasons that will be discussed below. Once realistic estimates of uncertainty are available, it should then be possible to rigorously validate the methods used to make sufficiently reliable in situ measurements.

This paper will not be specifically discussed ‘on site’ measurements, where a physical sample is removed, but the ex situ measurement is made near the original location of the test material, typically in a field laboratory. This class of measurement is clearly intermediate between those made in situ and ex situ. In terms of uncertainty estimation and method validation, ‘on site’ measurement can usually be treated as an ex situ measurement. This is because the primary sample that has been removed can be prepared and

Based upon the talk ‘The way forward for Uncertainty from Sampling’ presented by the author at the EURACHEM Workshop ‘Uncertainty from sampling and analysis for accredited laboratories’, November 2019, Berlin, Germany.

✉ Michael H. Ramsey
m.h.ramsey@sussex.ac.uk

¹ School of Life Sciences, University of Sussex,
Brighton BN1 9QG, UK

mixed in the field lab, so the uncertainty contribution from in situ heterogeneity contributes no more than it does for remote ex situ measurements.

In situ measurement methods

In situ methods are used in a wide range of different situations. These applications include a diverse range of analytes in environmental media, such as rocks, soils, waters and gases. They are increasingly also being used in the clinical sector, often at a patient's bedside at the 'point of care'. Such measurements are also widespread in the manufacturing industries, often to monitor product intermediates to optimise productivity. In situ chemical measurements are closely related to the wider concept of 'in situ testing' [3], which also includes many types of physical measurement.

Even though making an in situ measurement does not include the removal of a physical sample, there is a test portion of the test material which is 'interrogated' during the in situ measurement process. This 'undisturbed sample' has physical dimensions of volume and mass, which can be hard to estimate and, critically, may vary greatly among different analytes. For example, when making simultaneous in situ measurements of 19 elements by PXRF on pellets made of powdered silicates [4], the 'undisturbed sample' was estimated to have a mass ranging between 0.001 mg and 0.32 mg for Al and Ba respectively, for an 8 mm beam size.

Advantages of in situ measurement methods

The widespread adoption of in situ measurements largely arises from several inherent advantages. The measurement result is received promptly, virtually instantaneously, compared to days or weeks for those made in a distant laboratory. In many circumstances, this higher speed is advantageous, saving lives in the clinical sector or saving money in the commercial and environmental sectors. Another benefit of this immediacy is the possibility for the experimenter to follow up unexpected outcomes promptly. In this case, the results of the first round of measurements are used to immediately design and implement a second round whilst in the field, for example to investigate an anomaly discovered in the first round [5]. However, this approach should not be used to circumvent a systematic survey design, such as a predetermined grid, by introducing subjective judgement. A further benefit of the immediacy of results is the ability for a large number of these in situ devices (e.g. sensors), to be set up as a sensing networks to monitor analyte variability across either time or space, which is not otherwise readily feasible. One example of this is the deployment of a network of sensors in an urban environment to measure air quality [6].

The second advantage of in situ measurements is usually their substantially lower cost. This can have the obvious benefit of lowering the cost of an entire investigation, but can also be used to enable the taking of many more 'samples', an outcome that improves the appreciation of the entire area of study. For example, the lower cost of making in situ gamma-ray spectrometry measurements of ^{137}Cs at a former nuclear site enabled a more reliable survey for 'hot particles' than laboratory measurements made on much smaller sample masses taken with large gaps between them [7]. As is usual, the in situ measurements have higher analytical uncertainty, owing to the shorter counting time possible in the field. However, the overall uncertainty of the in situ measurement values is lower than for the ex situ, because the UfS is lower for the in situ method due to the much higher mass of the 'undisturbed' sample than for the removed sample measured ex situ (discussed in more detail under 'Treatment of bias'). Moreover, more 'samples' can be addressed by in situ measurement for the same expenditure, an important feature of surveys. It might well provide more information to make (say) 1000 less exact measurements than 100 more exact ones.

A third advantage is the ability to avoid sample preparation. For instance, the loss of some volatile analytes can occur by just removing the sample needed for laboratory measurement. The delay before a laboratory measurement can also cause further losses (e.g. of dissolved oxygen from water). Many of the traditional steps of sample preparation (drying, disaggregation, sieving, mixing, splitting and grinding of soils) can all cause further losses of analyte and also provide potential for contamination of the test material. Furthermore, some of these mechanical preparation steps can bias the analyte concentration. For example, sieving a soil at a 2-mm mesh size conforms to a widely used definition of 'soil' [8], but excludes larger particles and biota that would be included in an undisturbed sample. A closely related advantage is that they do not require a physical sample. That has the benefit of eliminating the costs of taking, storing and disposing of samples, which can be substantial in some sectors, such as the nuclear industry.

A fourth broad advantage of in situ measurements is that they can have the ability to quantify the heterogeneity of the test material, which provides extra valuable information not usually available to environmental scientists [9].

A supposed 'advantage', but really a disadvantage, of in situ measurements is that they can apparently be made by less skilled personnel than required for laboratory measurements. The person operating, or installing, the in situ measurement equipment actually needs to be *more* highly trained than is required in the laboratory. In situ measurement scientists need to be able to take the most appropriate samples (of the 'undisturbed' kind) in the real world, as well as taking measurements of acceptable quality, both without

local supervision. Disturbingly, because the quality of in situ measurements is not usually assessed rigorously, it is currently possible to use less skilled personnel without their adverse effects on data quality being detected.

Approaches to ensuring the quality of measurements, ex situ and in situ

Techniques for quantifying and improving the quality of ex situ measurements have been developed and well documented over the last 60 years. The most important and universal measure of data quality has proved to be the uncertainty. Informally, measurement uncertainty can be described as ‘the range of values within which the true value is asserted to lie’. Formally, measurement uncertainty is defined as a ‘parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand’ [10]. The ‘measurand’ is formally defined as ‘the quality quantity intended to be measured’, but is closely related to the traditional concept of ‘true value’ in statistical terminology, used for clarity in the informal definition. Uncertainty estimates, therefore, have both random and systematic components, usually quantified as some form of precision and bias, respectively. The procedures for estimating the measurement uncertainty that arises in the chemical laboratory are well established [11]. However, it has been recognised that the measurement process usually begins at the time that the primary sample is taken. Further guidance has therefore been made available on how to estimate the measurement uncertainty that includes that which arises from the sampling process [12]. The main method that has been adopted to estimate the random component of the uncertainty from sampling (U_{fS}) is the duplicate method. In this method, duplicate samples are taken at a small proportion of the sampling targets (e.g. 10 %, but no less than 8), selected at random to give a typical estimate of the U_{fS} for all comparable sampling targets. Both duplicated samples are also chemically analysed in duplicate, in a balanced design. The resultant measurement results (typically $32 = 8 \times 2 \times 2$) are evaluated statistically using analysis of variance (ANOVA) to give estimates of the measurement uncertainty, and its two components arising from the sampling (U_{fS}) and the chemical analysis.

Guidance on the validation of ex situ methods of chemical analysis, largely based upon the estimates of measurement uncertainty, has also been agreed internationally [13]. This guidance does not, however, include the wider definition of the measurement process to include the primary sampling step. Because the process of in situ measurement inherently includes the act of selecting and taking an ‘undisturbed

sample’, then the uncertainty of an in situ result has also to include uncertainty arising from sampling.

Ensuring the quality of in situ measurements

Despite the widespread adoption of in situ measurement methods, there is no agreement on how to ensure the quality of their measurement results. Measurement quality is considered even more important for in situ methods than for laboratory methods, because of the general lack of quality assurance and control (QA & QC), supervision and operator training in most fields of application. When important decisions are being made on the basis of in situ measurements, it is particularly important that a rigorous approach is devised for estimating measurement uncertainty. This will enable both the results to be interpreted probabilistically, hence more reliably, and also to allow in situ methods to be rigorously validated.

Currently, there is no universally agreed method for uncertainty estimation, and hence for validation, of methods of in situ measurement. Because of this absence, other approaches have been applied, often tending to be specific to a particular application sector. A general approach is often referred to as ‘Type testing’ by the manufacturers of the measurement equipment. Type testing has been defined in one sector as a ‘test performed to provide evidence that the design meets the requirements of the functional specification’ [14]. This ‘functional specification’ is defined as the ‘features, characteristics, process conditions, boundaries, and exclusions defining the performance of the tools’ [15]. One common physical example of the use of type testing is to the equipment used to measure the electricity consumed in a house. Once this sealed electricity meter is installed in a house, it is assumed that the results of the original test (usually carried out in the place of manufacture) are applicable to all of the subsequent measurements made over many years. The potential application of type testing to in situ chemical measurement is problematic in many respects. The analytes and the test materials are more diverse and the operating conditions are usually more variable and potentially damaging to the equipment. These conditions frequently lead to drift in the equipment’s performance over time. Furthermore, type testing ignores the U_{fS} due to heterogeneity of the analyte in the undisturbed sample.

There are several potential reasons why uncertainty estimation (and hence rigorous validation) of in situ measurement methods has not been widely undertaken. One reason includes the great diversity in the types of in situ methods, and the range of different application sectors, and hence physical situations, in which they are applied. For example, in the clinical sector there is not yet a wide acceptance of the concept of measurement uncertainty. Any approach to

method validation that relies upon the estimation of measurement uncertainty will currently not be readily understood or accepted, therefore, by front-line health care practitioners. A further reason is the widespread description of in situ measurement methods using terms such as ‘semiquantitative’, ‘rough and ready’, ‘screening’ or ‘indicative’. It is tacitly assumed that such methods do not require rigorous validation. A further impediment is the difficulty in assessing the validity of a method in the actual ‘field’ situation of application, as opposed to the place of manufacture. Ironically, the very absence across all application sectors of a widely accepted approach to uncertainty estimation, and thence validation, is the principal reason for not applying it.

One popular alternative option is the ‘validation’ of in situ measurement instrument by the manufacturer, usually in isolation and without a realistic sampling target. This can be done entirely on the manufacturer’s property, usually not in a ‘real world’ situation with a heterogeneous target, in an approach similar to the ‘type testing’ already discussed. The first limitation is that this approach can ignore the instrumental ‘drift’ that typically arises in the days, weeks and years after the manufacturer’s validation. Some manufacturers do address this by providing ‘check samples’ (or better, certified reference materials, CRMs) to monitor instrumental drift on a daily or hourly basis [16]. The use of CRMs to monitor and subsequently correct measurement made in situ has been recommended by some regulators [17]. One problem with this approach is that CRMs and check samples are often very different from the test material in the real world. For example, soil CRMs are often dry, finely ground, very homogeneous and often compacted into pellets with very little pore space or extraneous matter (e.g. biota in soil). By contrast, soils in the real world are usually moist, extremely variable in grain size, very heterogeneous and contain a high proportion of pore space and organic matter such as plant roots and animals. The systematic error (or measurement bias) that can occur from such causes can be revealed and quantified by comparing in situ measurements against ex situ measurements made by a different analytical method on the same sampling targets for the same measurand, as explained below. The heterogeneity of the analyte in the test portion for ex situ measurements is usually reduced to a minimum by grinding and mixing. In situ, however, this heterogeneity is often the biggest source of uncertainty, although it is usually ignored in procedures for both uncertainty estimation and validation.

Estimation of U for in situ measurements

There have been several publications describing possible approaches to estimating the uncertainty of in situ measurements, usually for the purpose of validation. These include

studies of metals in soil [1], bilirubin in infants [18], passive detection of ^{137}Cs by gamma-ray spectrometry [7] and oxygen isotope ratios in quartz by SIMS [2]. These studies will not be described in detail here, but only those aspects that illustrate general issues and help explain the outstanding challenges that need to be addressed.

The random components of the uncertainty of in situ measurements are usually estimated using the duplicate method, which has been widely employed for ex situ measurements, as already explained. The equivalent of the ‘duplicate samples’ is taken by placing the in situ measurement device twice, reflecting independent interpretations of the measurement protocol. For example, in the use of PXRF to measure Pb in topsoil [1] the instrument is placed on the soil at two estimates of the sampling location for a particular sampling target, separated by a distance representing the spatial uncertainty of the survey technique. In this example, the spatial uncertainty was ± 2 m, so the duplicate sample was located 2 m away from the first in a randomly chosen direction. These two sampling points are both equally likely interpretations of the protocol, given that particular surveying technology. The duplicate in situ readings will reflect the effect of the small-scale spatial heterogeneity, of the analyte concentration at that location, on the uncertainty. The use of at least eight such duplicate samples, selected at random across the investigation site, will reflect the typical measurement uncertainty caused by heterogeneity. When sampling in the temporal domain, for example for river waters, the investigator should take the duplicate samples with a time-lapse that similarly reflects the temporal ambiguity in the sampling protocol.

The measurement uncertainty estimated using the duplicate method alone does not include the systematic component arising from any bias in the chemical analysis or the field sampling. The bias from the chemical analysis alone is routinely estimated typically by measurements made on matrix-matched CRMs and can easily be included in the estimate of the measurement uncertainty (see Example A2, p 50 in [12]). The bias generated in the sampling process has proved much more difficult to evaluate and to consider in the uncertainty estimate. The approach most often adopted to estimate systematic sampling effects within in situ measurements, is to compare them against ex situ measurements made for the same measurand, on the same sampling targets. For the determination of Pb in top soil, take-away samples can be extracted at the locations where PXRF measurement was made and then prepared and analysed in a remote laboratory [1, 19]. The bias is then estimated as a function of concentration, typically by ordinary regression or, preferably, FREML (functional relationship estimation by maximum likelihood [20]), a technique in which the uncertainty of both variables is properly taken into account. Assuming no lack-of-fit, the slope of the linear model gives the

rotational component of the bias, and the intercept gives the translational component. An example of such a linear relationship for the Pb concentration [Pb] in top soil example (Fig. 1 in log space, where model appears non-linear) has an equation of

$$\text{In situ [Pb]} = 0.43(\pm 0.08) \times \text{Ex situ [Pb]} + 77(\pm 26) \quad (1)$$

This equation gives an estimated rotational bias of -57% ($\pm 8\%$) ($100 \times (1 - 0.43)$) of the in situ measurements by PXRF compared with ex situ measurements by AAS. Possible causes of the measurement bias were identified as soil moisture, material > 2 mm, surface roughness in the PXRF ‘undisturbed sample,’ and the difference in depth between the undisturbed sample for in situ PXRF (~ 1 mm) and the removed ex situ field sample for AAS (150 mm).

Challenges in uncertainty estimation for in situ measurements

Treatment of bias between in situ and ex situ measurements

One challenge is to decide what to do with this broader estimate of the measurement bias, including sampling bias. The most frequent action is to ‘correct’ the in situ measurements (using a regression-type model) to agree with the ex situ measurements. This procedure has been recommended in the environmental sector [17]. In this approach, it is also

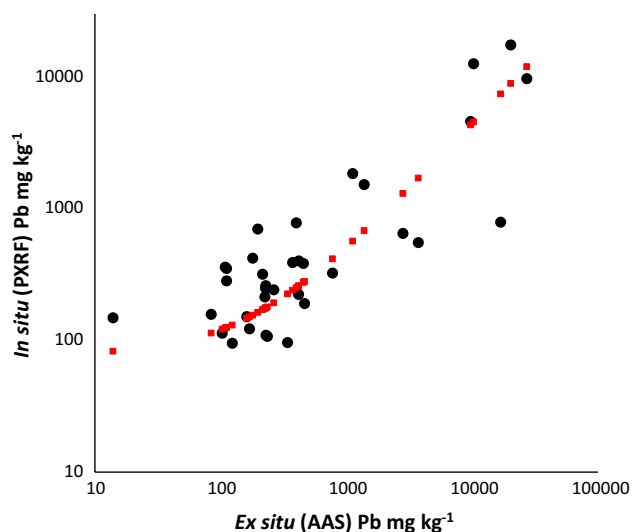


Fig. 1 Systematic component of uncertainty, or ‘measurement bias’, between in situ and ex situ measurements of Pb concentration (black circles) at 35 matching locations of top soil from one site, estimated using a linear FREML model (red squares Eq. 1), fitted to original data from [19]

possible to include the uncertainty of the estimated bias (e.g., $\pm 8\%$) into the estimate of the uncertainty of the corrected measurement value. One possible criticism of this approach is the implicit assumption that the ex situ measurements are closer to the true value of the analyte concentration (i.e., value of the measurand) [21]. In the other case study already discussed, (^{137}Cs in soils), the in situ gamma-ray measurements were shown to have a seven times lower uncertainty than the ex situ gamma-ray measurements [7]. This was primarily because the passive in situ measurements were based on a much larger mass of the undisturbed sample, estimated to be 150 kg compared to the 0.5 kg samples taken to the laboratory. In this case, therefore, it could be argued that the in situ measurements were more representative of the sampling targets, and therefore the ex situ measurement should be ‘corrected’ to agree with the in situ measurements. In this particular example, no significant bias was detected between the two sets of measurements, although this may be primarily due to the large uncertainty of the ex situ measurements (73 %).

An alternative approach to this issue is to not ‘correct’ either set of measurements, but only to include any possible systematic effects into the estimate of the uncertainty of each measurement. In the instance of ^{137}Cs , the large ex situ measurement uncertainty ($U=73\%$) caused by the low mass of the primary sample (0.5 kg) tells us that the true value may be much further away from the measured value than observed for the in situ measurements ($U=13\%$). From a purely metrological point of view, this method of comparison could be criticised because the ‘test portion’ masses are not matched. However, both types of sample are intended to represent to same sampling target, which is defined as a ‘portion of material, at a particular time, that the sample is intended to represent’ [10]. In this example, the mass of the sampling target was matched in both cases, because the samples and measurement were taken on the same sampling grid for the same purpose. Both the in situ and the ex situ sample at the location were, therefore, intended to represent the same sampling target, which was centred on the same point on the sampling grid.

Selection of reference or ex situ method to estimate bias

The second challenge is finding a suitable measurement method (e.g. ex situ) against which to estimate measurement bias. This can be illustrated using an example from the clinical sector, namely the measurement of transcutaneous bilirubin (TcB) in the tissue of the new-born infant, which can be used to decide whether the infant needs treatment for jaundice [18]. In this procedure, a hand-held reflectance photometer is placed on infant’s forehead in three places to make a composite measurement of absorbance at 450–550 nm. The

result is used to decide whether a more invasive blood sample need be taken from the heel of the infant, and an ex situ measurement made of the total serum bilirubin (TSB). The ex situ TSB measurements are made using direct absorbance spectrophotometry at 455 nm on centrifuged blood samples and can also be used to estimate the bias. Interestingly, the potential heterogeneity of the optical density in the baby's skin is implicitly recognised by the use of the three-fold composite to reduce its effect. On site checking for drift in the reflectance photometer calibration is made using a 'checker plate' with two inspecting screens, one white and the other yellow. For TSB the random component of the uncertainty, here called within-run precision, was estimated by using fivefold replicated measurements. The estimated coefficient of variation varied from 0.5 to 0.9 % as the concentration of TSB increased. However, the more important precision estimates for the TcB measurements themselves were not reported for unexplained reasons. The systematic component of the uncertainty of this in situ method, called accuracy by these authors, was estimated by comparing the in situ TcB measurements against ex situ TSB measurements made on the same matching 271 infants (Fig. 2). The model of the relationship, between the two methods, made using Deming regression, gave the equation:

$$\text{TcB} = -5.5012 + 1.0160 \text{TSB}.$$

(Note: Deming regression is similar to FREML but based on least-squares estimation.)

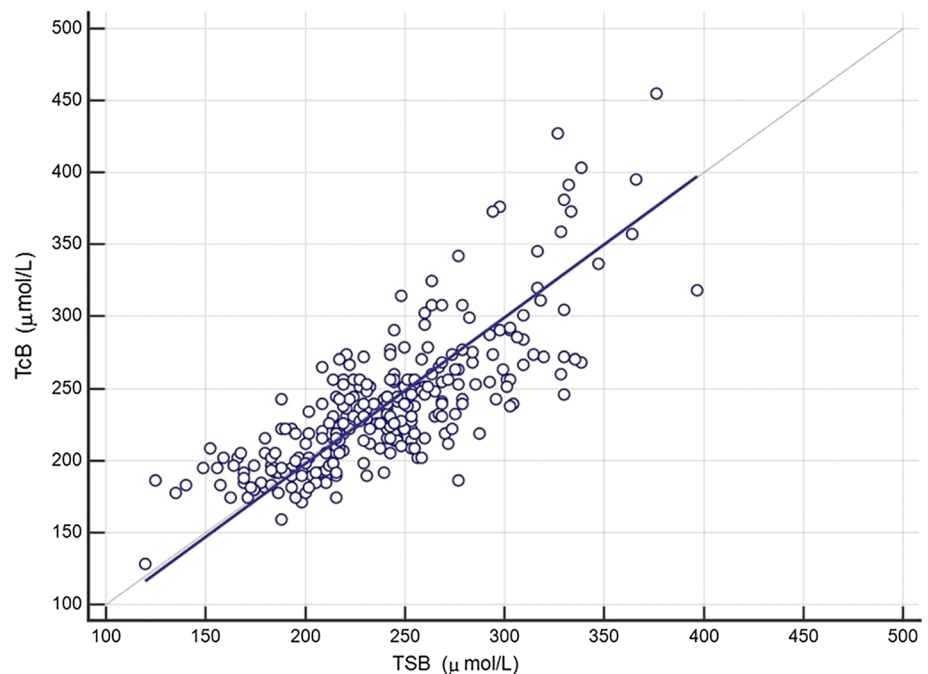
No significant bias was detected between the two methods, because the slope coefficient did not differ significantly from unity (95 % CI 0.87 to 1.16) and the intercept

coefficient did not differ significantly from zero (95 % CI -39.71 to 28.71). Despite no detection of significant bias, the authors reported that 'TcB...tended to underestimate the TSB, but measurements of TcB could underestimate or overestimate TSB values at both low and high bilirubin levels.' This comment suggests a nonlinear relationship, and hence use of an inappropriate linear regression model. This may be one reason that the authors recommended that different threshold values be used as the upper safe limit for TcB measurements ($222 \mu\text{mol/L}$) compared to that for TSB measurements ($291 \mu\text{mol/L}$).

The generally accepted approach of measuring CRMs to validate either the TcB or the TSB methods was not used, for unexplained reasons. This approach would have been useful, for example in choosing between several different ex situ methods, which the authors report as not even being consistent between themselves. The concept of measurement uncertainty would have allowed the inclusion of both the systematic and random components in all of the measurements, and hence a more rigorous comparison and validation of both the in situ and the ex situ methods.

It is interesting that there was not a perfect match between the measurands specified for the two methods. This appears to be a potential problem for in situ measurements in general. The effective measurand for the TSB method was the *intravascular* bilirubin concentration, while for the TcB it was the *extravascular* bilirubin concentration. The authors state, therefore, that the 'TcB should not be expected to equal the TSB when comparing the two bilirubin measurement methods'. One possible approach in this example is to consider the in situ TcB measurements as merely proxies

Fig. 2 Fitted line showing regression model of the relationship between in situ measurements of transcutaneous bilirubin (TcB) against ex situ measurements of total serum bilirubin (TSB) on 271 infants. (Taken from [18] with permission of Elsevier)



for the ex situ TSB measurements upon which the clinical decisions are then made. However, measurements from both methods seem to be used for clinical decisions. In terms of matching the sampling targets, it could be specified as matching for both methods, as ‘the particular infant that may have jaundice’. However, at the more specific level, the target differs, and the main clinical diagnosis is largely based upon the bilirubin concentration in the *serum* (TSB), whilst the in situ TcB device measures the bilirubin in the *tissue* (i.e. skin). The sampling media between the two methods are evidently not matched, even though the broader sampling targets are.

For comparison, a very similar but independent study of TcB versus TSB [22], also reported evidence of some non-correspondence between the two sets of measurements, with measured TcB generally greater than TSB at low values of TSB, and less than TSB at high values of TSB (i.e. > 16 mg/dl). Again, the precision of the TcB measurements was not reported, and CRMs were not used to estimate bias or traceability.

This example shows that finding a matched method to estimate the systematic component of the uncertainty of in situ measurements cannot always be achieved perfectly. One solution would be to explain, with the reported uncertainty value in the validation, the degree to which the measurand and the sampling target of the two methods are matched.

Adequate staff training and supervision

The third general challenge for in situ methods is the need for adequate education, training and supervision. At a fundamental level, there needs to be an understanding of the concept of measurement uncertainty in all of the different application sectors and by all practitioners. For example, in the clinical sector the focus is still on the precision and bias of analytical methods, rather than uncertainty of the resultant measurement values. At the level of the operator of the in situ measurement device, training and supervision is also a key challenge. In the specific bilirubin example, it was noted that there was a worse correlation between TcB and TSB measurements when the infants were measured as outpatients rather than as inpatients [18]. This provides evidence for the need for appropriate training of the operators, particularly when they make the measurements away from the more supervised environment of the laboratory (or hospital). As noted previously, the apparent ease with which many in situ measurements can be made belies the higher level of skill and training that is required to make reliable measurements ‘in the field’ (e.g. at the point of care), than in the laboratory with its more embedded QC and QA systems. Tools to assess the quality of in situ measurements, and hence effectiveness of the training, are already being

applied in some sectors. These tools include internal quality control (IQC), in which the operator routinely monitors the ongoing quality of the in situ measurements using both CRMs (or check materials) and a small proportion of duplicated measurements. An even more powerful tool is to participate in External Quality Assurance procedures (EQA), such as proficiency testing (PT), which has already been implemented for in situ stack gas emission measurements [23]. The PT results can not only demonstrate the quality that operators achieved in routine operation, but also provide feedback to the operators and their supervisors of their performance and the benefits of training. Adoption of these QC/QA tools is needed urgently for in situ measurements, because of the greater skill and training needed, and lack of supervision, in the difficult locations where most operators often work.

General applicability of the uncertainty estimates

The fourth challenge for in situ methods is to test the applicability of reported uncertainty estimates. Clearly the effect of the analyte heterogeneity in the sampling target needs to be included in estimate of the uncertainty of measurements made by the in situ device. If the level of heterogeneity is similar for most sampling targets, then it can be assumed that the uncertainty estimate made at the time of validation will be broadly applicable for all targets. In the bilirubin example, if the variability of TcB within all young infants’ foreheads is reasonably constant, then the uncertainty estimate made at the time of validation will be broadly applicable. Traditional quality control practice would then be to routinely duplicate a small proportion of routine TcB triplicate measurements (e.g. 10 %, with a second set of three new measurement points per infant). A control chart could then be used to monitor whether the general level of TcB variability conforms to that found during the initial validation. By contrast, for the example of Pb in topsoil, the Pb concentration (as mass fraction) can vary widely from 10 to 30000 mg/kg and the heterogeneity (quantified as U_{HET} using the same in situ method) can also vary widely. When heterogeneity is expressed as relative expanded uncertainty of U_{HET} , it is very variable between sites, ranging from 1 to 100 % [9]. Across twelve different sites, heterogeneity was not found to increase as a function of Pb concentration, but rather as a function of spatial scale within some of the sites. In this situation, an estimate of measurement uncertainty for Pb concentration in soil by PXRF made at one site would not necessarily be applicable to such measurements made on soils at another sites.

Three options for overcoming this fourth challenge are possible. For ex situ measurements made in the laboratory, methods are validated for a limited range of composition. For example, an ICP-AES calibration might be designed and

validated for lightly contaminated loamy soils (composed mainly of silicate minerals, with $\text{Pb} < 500 \text{ mg/kg}$) but this validation would not be applicable to measurements made on soils that have extremely different overall composition (e.g. very calcareous, where Ca causes substantial matrix effects on many analytes), or heavily contaminated with $\text{Pb} > 5000 \text{ mg/kg}$. Similarly, for in situ measurements of Pb in soil, an uncertainty estimates made at this validation site would not automatically be considered applicable to all sites. Validation (and uncertainty estimates) would need to be quoted for test materials within a specified range of chemical composition.

A second option would be to use a general model of UfS (expressed as s_{sam}) versus analyte concentration (c) to predict the UfS, and hence the overall measurement uncertainty. Such a model has been described in the food sector in a meta-analysis of results for 75 analytes in 27 types of food, from field, store, factory to retail outlets [24]. The relationship discovered is shown as a power law in Fig. 3, and the equation of the fitted model was:

$$s_{\text{sam}} = 0.0128c^{0.8099}.$$

Predictions from this model have relatively large confidence interval, as the residual standard deviation was 0.46, so the factor applying to predicted s_{sam} is $10^{0.462 \times 2}$, which is almost an order of magnitude. However, the model might still make a useful initial estimate of UfS, and hence of measurement uncertainty. Most of the case studies used in the meta-analysis used traditional ex situ sampling and measurement, so the validity of the model

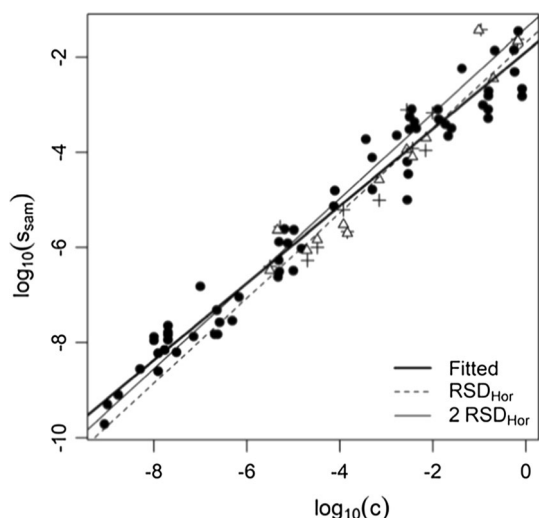


Fig. 3 The linear relationship between the logarithms of the uncertainty from sampling (expressed as s_{sam}) and the analyte concentration (c), based upon a meta-analysis across the food sector, compared to the Horwitz function (RSD_{Hor}). (Reproduced from [24] with permission from The Royal Society of Chemistry)

for in situ measurements, with their ‘undisturbed’ samples, would need to be separately tested. Such studies need to be repeated, with both in situ and ex situ measurements, and also in other application sectors (such as soils, waters and gases). The first reason for this further meta-analysis is to be able to test whether the equation found in the food sector has broader applicability (rather like the Horwitz function for different analytical systems, Fig. 3). Secondly, equations in each different sector, even if dissimilar, could be used to make initial estimates of UfS and measurement uncertainty (e.g. for regulators) in particular application sectors.

The third option would be to treat the two main sources of uncertainty separately. The measurement uncertainty estimate for the instrument (U_{inst}) could be estimated in isolation, as is often done currently, with an effectively homogeneous test material. The uncertainty due to the heterogeneity of the analyte within the sampling target (U_{HET} , equivalent to UfS), as specified in the measurement protocol for that test material, would need to be estimated as a separate objective. This could be achieved using the duplicate method, applying the same in situ measurement device, on an appropriate number of sampling targets (e.g. at least 8), that are typical of that applications sector. The two components of the uncertainty could then be combined together to give an estimate of the measurement uncertainty using the equation:

$$U_{\text{meas}} = \sqrt{U_{\text{inst}}^2 + U_{\text{HET}}^2}$$

If the range of concentration was relatively small and well above the detection limit, then this equation might be sufficient. However, if the range was large and in close proximity to the detection limit, then the summation would have to express U as a function of concentration for all three terms in the equation. Alternatively, a hybrid option could be to use predictions of UfS of in situ measurements from the meta-analysis model (from Option 2) for the initial estimate of U_{HET} in Option 3.

For test materials where misclassification would have high financial consequences, it may be worthwhile to make an estimate of U_{HET} that is specific to that site, say with at least 8 duplicated samples taken across the site. Balancing the measurement uncertainty against the cost of misclassification caused by excessive uncertainty can be made using the optimised uncertainty procedure [1].

Similarly, as part of IQC for the bilirubin example, a small proportion of duplicated threefold measurements of TcB could be made on the forehead of randomly selected infants, to check whether they have a significantly higher level of analyte heterogeneity (as U_{HET} and hence UfS) than the value assumed in the model.

Action required

To agree a universally applicable procedure to estimate the uncertainty of in situ measurements, there needs to be a meeting of experts from a wide range of application areas. This agreed procedure could then be used as the bases for a second universally accepted procedure for the validation for these methods. A third requirement is to agree the quality control procedures needed to assess the quality of the routine in situ measurements, when they are made by operators in the real world with all of its limitations of time and resources.

Conclusions

Chemical measurements made in situ have many advantages over those made in a remote laboratory, but they do tend to have higher levels of measurement uncertainty even when conducted competently. An estimate of this uncertainty needs to be made in order to make sure that the in situ measurement results are fit-for-purpose (FFP), and hence to make a reliable interpretation of the results. The main problem lies in ensuring that the measurements are indeed conducted competently.

Both ex situ and in situ measurement processes include the sampling step, but for in situ measurements sampling per se is often not recognised, as the ‘undisturbed sample’ is usually left in place and is of unknown dimensions. Estimates of the uncertainty of in situ measurement values should never be based solely upon the instrumental repeatability, as this will always lead to an underestimation. The heterogeneity of the analyte concentration within the sampling target, and also within the undisturbed sample taken to represent it, causes extra uncertainty in the measurement value. This is effectively the same concept as the uncertainty from sampling (UfS) that is associated with taking a physical primary sample for traditional laboratory measurement. Validation of all analytical procedures needs to use realistic estimates of measurement uncertainty. These estimates need to include UfS, particularly for in situ measurements where the sampling step is often not recognised. Four main challenges for estimating the uncertainty of in situ measurements have been identified, and some possible solutions described: (1) treatment of bias between in situ and ex situ measurements, (2) selection of a reference or ex situ method to estimate bias (3) adequate training and supervision, and (4) generally applicable uncertainty estimates.

Further consultation and discussions are urgently required between specialists from the many different

sectors where in situ methods are applied, to agree three universal procedures for in situ measurements, to (1) estimate the uncertainty (2) validated the methods, and (3) implement quality control to monitor in situ measurement quality in routine operation.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Taylor PD, Ramsey MH, Potts PJ (2004) Balancing measurement uncertainty against financial benefits: a comparison of in situ and ex situ analysis of contaminated land. *Environ Sci Technol* 38:6824–6831
2. Ramsey MH, Wiedenbeck M (2017) Quantifying Isotopic Heterogeneity of Candidate Reference Materials at the Picogram Sampling Scale. *Geostand Geoanal Res* 42(1):5–24
3. <https://www.sciencedirect.com/topics/engineering/insitu-testing>. Accessed 20 Jan 2020
4. Rostron P, Ramsey MH (2017) Quantifying heterogeneity of small test portion masses of geological reference materials by PXRF: implications for uncertainty of reference values. *Geostand Geoanal Res* 41(3):359–473. <https://doi.org/10.1111/ggr.12162>
5. Potts PJ, Ramsey MH (2002) Carlisle J (2002) Portable X-ray fluorescence in the characterisation of arsenic contamination associated with industrial buildings at a heritage arsenic works site near Redruth, Cornwall, UK. *J Environ Monit* 4:1017–1024
6. Blythe, PT, Neasham J, Sharif B, Watson P, Bell MC, Edwards S, Suresh V, Wagner J, Bryan H (2008) An environmental sensor system for pervasively monitoring road networks. In: IET—road transport information and control 2008. IEEE, Manchester, pp 1–6
7. Rostron PD, Heathcote JA, Ramsey MH (2014) Comparison between in situ and ex situ gamma measurements on land areas within a decommissioning nuclear site: a case study at Dounreay. *J Radiol Prot* 34:495–508
8. MAFF (Ministry of Agriculture, Fisheries and Food) (1986) The Analysis of Agricultural Materials: A Manual of the Analytical Methods Used by the Agricultural Development and Advisory Service, 3rd edn. Ministry of Agriculture, Fisheries and Food, HMSO, London
9. Ramsey MH, Grace Solomon-Wisdom GO, Argyraki A (2013) Evaluation of in situ heterogeneity of elements in solids: implications for analytical geochemistry. *Geostand Geoanal Res* 37(4):379–391. <https://doi.org/10.1111/j.1751-908X.2013.00236.x>
10. JCGM 100:2008 Evaluation of measurement data - Guide to the Expression of Uncertainty in Measurement (GUM). Sevres, <http://www.bipm.org/en/publications/guides/gum.html>. Also available as ISO/IEC CD Guide 98-3, Evaluation of measurement

- data—Part 3: Guide to uncertainty in measurement (GUM), International Organisation for Standardisation, Geneva
11. Ellison SLR, Williams A (eds) (2012) Eurachem/CITAC guide: quantifying uncertainty in analytical measurement, 3rd edn. ISBN 978-0-948926-30-3. <http://www.eurachem.org/>
 12. Ramsey MH, Ellison SLR, Rostron P (eds) (2019) Eurachem/EUROLAB/CITAC/Nordtest/AMC guide: measurement uncertainty arising from sampling: a guide to methods and approach, 2nd edn. Eurachem, ISBN 978-0-948926-35-8. <http://www.eurachem.org/index.php/publications/guides/musamp>
 13. Magnusson B, Örnemark U (eds) (2014) Eurachem guide: the fitness for purpose of analytical methods—a laboratory guide to method validation and related topics, 2nd edn. ISBN 978-91-87461-59-0. www.eurachem.org
 14. ISO 13880:1999, Petroleum and natural gas industries—content and drafting of a technical specification, 1st edn (1999). Global Standards. <https://www.iadclexicon.org/type-test/>
 15. API SPECIFICATION 19TT, Specification for Downhole Well Test Tools and Related Equipment, First Edition, October 2016. Global Standards
 16. <https://www.nitonuk.co.uk/wp-content/uploads/NitonUK-XL3t-GOLDD-XRF-Analyser-V1-Sep-19.pdf>. Accessed 20 Jan 2020
 17. Environment Agency (2009) Framework for the use of rapid measurement techniques (RMT) in the risk management of land contamination. ISBN: 978-1-84432-982-3
 18. Ercan S, Özgün G (2018) The accuracy of transcutaneous bilirubinometer measurements to identify the hyperbilirubinemia in outpatient newborn population. *Clin Biochem* 55(2018):69–74
 19. Taylor P (2003) Uncertainty of geochemical measurements of contaminated land: causes, estimation and cost-based optimisation. D Phil. Thesis, University of Sussex, UK
 20. Analytical Methods Committee (2002) Fitting a linear functional relationship to data with error on both variables, Technical Brief No.10, Royal Society of Chemistry, London. <http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp>
 21. Ramsey MH, Boon KA (2012) Can in situ geochemical measurements be more fit-for-purpose than those made ex situ? *Appl Geochem* 27:969–976. <https://doi.org/10.1016/j.apgeochem.2011.05.022>
 22. Maisels MJ, Engle WD, Wainer S, Jackson GL, McManus S, Artinian F (2011) Transcutaneous bilirubin levels in an outpatient and office population. *J Perinatol* 31:621–624
 23. Coleman MD, Robinson RA, Williams MB, Clack MJ, Butterfield DM (2013) State of UK emissions monitoring of stacks and flues: an evaluation of proficiency testing data for SO₂, NO and particulates. *Accred Qual Assur* 18:517–524. <https://doi.org/10.1007/s00769-013-1011-x>
 24. Ellison SLR, Ramsey MH, Lawrance P, Stuart B, Minguez J, Walker MJ (2017) Is measurement uncertainty from sampling related to analyte concentration? *Anal Methods* 9:5989–5996. <https://doi.org/10.1039/C7AY00752C>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.